An Overview of Functional Linear Models

Pang Du

Department of Statistics
Virginia Tech

Functional linear models

A linear model can become functional in one or both of two ways:

- The response variable y with argument t is functional.
- One or more of the covariates x is functional.

FLM: Some scenarios

- Chapter 13: functional response with categorical covariates (*functional analysis of variance*) or scalar covariates (*functional multiple regression*).
- Chapter 14: functional response depends on functional covariates concurrently (*concurrent model*).
- Chapter 15: scalar response and functional covariates.

The average Canadian weather data

- 35 Canadian weather stations selected to cover the country.
- Daily temperatures (0.1 degrees Celsius) and precipitations (0.1 mm) averaged over the years 1960 to 1994. (Feb 29th combined with Feb. 28th).
- Canada divided into Atlantic, Continental, Pacific and Arctic weather zones.

35 Canadian weather stations

Variables in Canadian weather data

- Response variable: a functional response Prec(t) measuring average precipitation over time t or a scalar response Prectot measuring the total yearly precipitation averaged over the years.
- Categorical covariate: climate groups referring to the Atlantic, Continental, Pacific and Arctic weather zones.
- Functional covariate: average temperature Temp(t) measured over time t.

Problem setup

- Scientific question: How does the shape of the mean annual precipitation profile depend on which climate zone the station is in?
- Climate zones indexed by $g = 1, 2, 3, 4$.
- Weather station in each zone indexed by $m = 1, \ldots, N_g$.

Model

The model for the mth precipitation function in the gth zone is

$$\text{Prec}_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t).$$

- $\mu(t)$: the grand mean profile across all 35 weather stations
- $\alpha_g(t)$: the *effect functions* representing departures from the grand mean specific to climate zones.
- $\epsilon_{mg}(t)$: residual effect at the weather station.
- Zero sum constraint:

$$\sum_g \alpha_g(t) = 0 \text{ for all } t.$$

Some interesting questions

- *Functional contrasts*: "there are no differences in mid-summer between the climate zones"
- Multiple comparisons: "Over which time intervals are there significant differences between climate zones?"

Problem setup

- Scientific question: Does the total amount of precipitation depend on specific features of the temperature profile of a weather station?
- Response variable:

$$\text{Prectot}_i = \int_0^{365} \text{Prec}_i(t)dt,$$

for $i = 1, \ldots, 35$.

Model

The model for the *i*th total precipitation is

$$\text{Prectot}_i = \alpha + \int_0^{365} \text{Temp}_i(t)\beta(t)dt + \epsilon_i.$$

- α: the intercept adjusting for the origin of the precipitation variable.
- $\beta(t)$: the regression coefficient function and the functional parameter of interest.
- ϵ_i: residuals.

Comparison with multiple linear regression

- Multiple linear regression:
  - n observations: $(y_i, \mathbf{x}_i), i = 1, \ldots, n$
  - each $\mathbf{x}_i$ is a p-vector: $\mathbf{x}_i = (x_{1i}, \ldots, x_{pi})'$.
  - the p-vector parameter $\boldsymbol{\beta}$ of coefficients is estimable if $n > p$.
- Functional linear regression:
  - n observations: $(y_i, x_i), i = 1, \ldots, n$
  - each x_i is a function with dimension$= \infty$.
  - how to make sure the coefficient function $\beta(t)$ is estimable?

A FPCA solution

- Karhunen-Loéve decomposition of the covariance function $K(s,t)$ of $x(t)$:

$$K(s,t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t),$$

  where $\{\lambda_k, e_k(t)\}$ are eigenvalues and eigenfunctions of the linear operator $f(t) \mapsto \int K(s,t) f(s) ds$ associated with K.
- $x(t)$ can be expressed as $x(t) = \sum_{k=1}^{\infty} \theta_k e_k(t)$, where $\theta_k = \int x e_k$.

A FPCA solution: The truncation

- **truncate** the series at a small K: so $x(t)$ is approximated by $x(t) \approx \sum_{k=1}^{K} \theta_k e_k(t)$
- expand $\beta(t)$ using the same truncated basis functions: $\beta(t) = \sum_{k=1}^{K} \beta_k e_k(t)$.
- essentially a multiple linear regression problem again.
- Drawback: the basis system $\{e_k(\cdot) : k = 1, \ldots, K\}$ may not be an effective one for representing $\beta(t)$.

A regularization solution

- add a roughness penalty on $\beta(t)$,
- essentially cut down the dimension of β by enforcing certain smoothness constraint on β.

Problem setup

- Scientific question: How does a precipitation profile depend on the associated temperature profile?
- There are several versions of models that are relevant under different assumptions on the influence range of the temperature profile.

The concurrent model

- Assumption: precipitation at time t depends only on the temperature at time t.
- The *concurrent model* for the ith precipitation function is

$$\text{Prec}_i(t) = \alpha(t) + \text{Temp}_i(t)\beta(t) + \epsilon_i(t).$$

- A special case of the *varying coefficient models* in Hastie and Tibshirani (1993).
- Should we use regularization to force β to be smooth in t?

The annual or total model

- Assumption: temperature influence on precipitation Prec(t) at a fixed time point t extends over the whole year.
- The annual model for the ith precipitation function is

$$\mathrm{Prec}_i(t) = \alpha(t) + \int_0^{365} \mathrm{Temp}_i(s)\beta(s,t)ds + \epsilon_i(t).$$

- The value $\beta(s,t)$ determines the impact of temperature at time s on precipitation at time t.
- We need roughness penalties for variation in both s and t.

The short-term feed-forward model

- Assumption: precipitation Prec(t) at a fixed time point t only depends on temperatures over a small time interval back.
- Example: falling temperature before precipitation.
- The corresponding model for the ith precipitation function is

$$\text{Prec}_i(t) = \alpha(t) + \int_{t-\delta}^{t} \text{Temp}_i(s)\beta(s,t)ds + \epsilon_i(t).$$

- δ: a pre-specified time lag over which we use temperature information.
- $\beta(s,t)$ is only defined on a complicated trapezoidal domain: $t \in [0,365], t-\delta \le s \le t$.
- More care taken for non-periodic data on the boundary of the domain.

The local influence model

- Assumption: precipitation Prec(t) at a fixed time point t only depends on temperatures over a t-dependent set Ω_t.
- Example: thunderstorm in summer comes with falling temperature before it and rising temperature afterwards.
- The corresponding model for the ith precipitation function is

$$\text{Prec}_i(t) = \alpha(t) + \int_{\Omega_t} \text{Temp}_i(s)\beta(s,t)ds + \epsilon_i(t).$$